Multidimensional scaling algorithm and its current applications in wireless localization

Kulvir Kaur, Student (M.Tech)
Electronics and communication engg. Punjabi university
Patiala, Punjab, India
Shergill_14@yahoo.co.in
Manjit Sandhu,Sr.Lecturer
Electronics and communication engg., sbbsiet
Jalandhar, Punjab, India
Manjit_sandhu67@yahoo.com

***Abstract*— The paper gives an overview of multidimensional scaling (MDS) algorithm. Aside from brief introduction to MDS, its characteristics, its various types and its current applications has been discussed in wireless systems i.e. localization of wireless sensors in a network and determining the position of a mobile station without GPS. MDS is an algorithm by means of which information contained in a set of data is represented by a set of points in a low dimensional space. These points are arranged in such a way that distances between the points represents the relation between the data elements.**

***Keywords*—MDS algorithm, MDS types, MDS current applications, localization**

I. Introduction

MDS is a set of data analyses technique. Information contained in a set of data is represented by a points in a multidimensional space such that the distance between the points tells about the relation between data sets i.e. their similarity or dissimilarity. Two data objects, if they are similar are represented by points that are close to each other in space and dissimilar objects are represented by points that are far apart.

A. Example of MDS

Suppose in a survey, it has been asked to compare the legal offences in term of their seriousness. The resulting data is represented in the form of a table 1. Each entry in the table tells the percentage by which respondents give their views of dissimilarity between offences.

The given information can be represented easily by MDS analyses and represented the data in 2D space as shown in fig1

TABLE 1
RESPONDANTS VIEWS

	1	2	3	4	5
Assault	0				
Embezzlement	21.1	0			
Perjury	71.2	54.1	0		
Libel	36.4	36.8	37.1	0	
Burglary	52.1	54.1	52.1	0.7	0
Stolen gold	89.9	74.2	36.5	54.1	53.0

Figure 1: 2D scaling of data of Table 1.

In the fig.1, the distance between point 4 and 5 is small but 1 and 6 is large which shows that assault and receiving stolen gold are very unalike but burglary and libel are little alike according to respondents views.

B. Literature survey

MDS is a set of data analysis methods, which allow one to infer the dimensions of the perceptual space of subjects. The primary outcome of an MDS analysis is a spatial configuration, in which the objects are represented as points in such a way, that their distances correspond to the similarities of the objects. Their detailed study has been given in [1],[2],[3]. First well known MDS proposal was given by torgerson in 1952 that fits Euclidean distance model. It was based on metric analyses i.e. quantitative approach. In 1950, Attneave gives a Non-Euclidean distance model but it was not so successful. In 1962, Shepard gives psychometric approach. In 1964, Kruskal gives a non metric approach i.e. based on qualitative analyses. Till then new researches are going on and various softwares have also been introduced for MDS like INDSCAL, ALSCAL etc. MDS has been used in various applications. Main is used in wireless localization like sensor network localization and mobile station localization.

A wireless sensor network [5] comprises of a large number of wireless sensor nodes and is randomly deployed. The location information does not only expand application scenarios, but also enhances performance of network protocols. In many applications, such as environmental monitoring, disaster detection, and battlefield monitoring, location information is vital. Deployments of wireless sensor net-works can be roughly categorized into manual settings, GPS-based approaches, and localization algorithms. It is not

DOI: 01.IJCSI.02.01.42

practical to manually set sensor location, because it is too time-consuming and expensive.

For the GPS-based localization method, the current/ existing GPS component is too big to fit into a tiny wireless sensor node. Another drawback is that GPS cannot be used in indoors. Locations of sensor nodes can be computed with the localization algorithm. A great deal of work over the past few years has focused on the sensor localization problem. A detailed survey of this area is provided in [6]. In the case when all pair-wise distances are known, the coordinates can be derived by using a classical method known as multidimensional scaling (MDS) [7], [8]. Shang et al. proposed the MDS-MAP localization algorithm using multidimensional scaling. MDS-MAP introduced in [9] can be mentioned as a well known example of this class where it computes the shortest paths between all pairs of nodes in order to approximate the missing entries of the distance matrix. Another algorithms that mainly consider the sensor localization as a non-convex optimization problem and directly estimate the coordinates of sensors is a relaxation to semi-definite programming (SDP) [10]. In [11] where they use matrix completion methods as a mean to reconstruct the distance matrix. Their analysis is based on a number of strong assumptions. First, they assume that even far away sensors have non-zero probability of detecting their distances. Second, the algorithm explicitly requires the knowledge of detection probabilities between all pairs. MDS MAP require weaker assumptions. It use shortest paths for the missing entries in the distance matrix and apply MDS to find the topology of the network. Here it is assumed that only neighbouring sensors have information about each other and that only connectivity information is known and the knowledge of detection probabilities plays no role in this algorithm. MDS-MAP still has the following drawbacks: 1) The time complexity is high. When there are a huge number of sensors, large bandwidth and computation are required to estimate locations. 2) In the case of a network area with non-convex shape, the computed coordinates are often erroneous. To resolve the drawbacks of MDS-MAP, a HMDS [12] is proposed which has low time complexity and which can operate in non-convex network environments.

Mobile locations could support many applications, such as emergency services, roadside assistance, navigation, tracking, and so on [13]. With the popularization of wireless services, more and more people call for emergency services by wireless phone, and the system operators want to provide subscriber's location information which conforms to the FCC requirement. Location estimations can be done through many ways. The first is the Global Positioning System (GPS) [14], which provides accurate positioning but fails when satellite signals are blocked like indoor situations. The second is MS location based on RSS measurements [15]. This method is has to measure the power of the signal at the receiver and low accuracy is achieved in this way due to the complex propagation mechanisms. The third way uses the angle-of-arrival (AOA) [16]. However, this method requires a complex antenna array system and suffers from NLOS propagation. The fourth method is the TOA [17]. NLOS propagation is a potential disadvantage of TOA method. The MDS has been proposed for wireless network based mobile location. Due to the MDS needs distance measurement data between the MS and BSs, the MDS technique has been proposed for mobile location using the TOA method [18]. It improves the performance of mobile location estimation.

C. Algorithmic flow of MDS

It is shown in fig. 2. Proximities are the data for MDS analyses. It can show similarity vs. A-similarity and can be symmetric vs. Asymmetric. MDS model can be metric/non metric with Euclidean distance or non Euclidean distance. Dimensions can be 1,2or more depending on the parameters that are to be surveyed. Analyses can be individual, aggregated or weighted.

Details are given ahead. Various softwares are used for MDS analyses that gives the final results of coordinates of data in the space. Most commonly used is ALSCAL developed by Forrest Young.

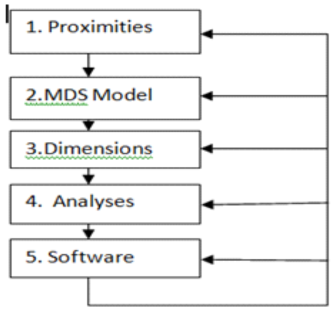

Figure 2: Flow for MDS analyses

II. CHARACTERISTICS OF MDS

There are four main characteristics that a researcher must consider while using MDS:-

- *The data :* It is simply the data that is to be presented for MDS analyses. It can be in the form of matrix. Distinction between data can be made on 'way of data' and 'modes of data'.
- *Way of data :* It simply represent the dimensionality matrix like 'One way data' It is like a single column matrix that can't be scaled i.e. can't apply MDS. 'Two way data' It is represented in matrix form and can be scaled by MDS.
- *Modes of data :* It tells about the kind of matrix the input data has like 'One mode data' has rows and columns refer to the same set of objects. The matrix is square and symmetric. 'Two mode data' has rows and columns of matrix refers to different set of objects. The matrix is rectangular and asymmetric.

The transformation technique : It rescales data in a matrix into an n –dimensional space. Two types:-monotonic and linear. 'Monotonic' preserves ordinal information of the data. It assumes rank order of entries in data matrix contain significant information. 'Linear' are expressible in simple mathematical form and are simpler. Data dissimilarities are direct estimate of distance between the points.

$$d_{jk} = b\delta_{jk}.$$

DOI: 01.IJCSI.02.01.42

- *The models* The model represents data value as distance between the points in n- dimensional space. Main model used in MDS is 'Euclidean distance model' the data in the matrix is represented as 'distance like'. It can be unweighted and weighted. In 'Unweighted model',all dimensions are given equal importance. Distance between points i and j in dimension a is given by:

$$d_{ij} = \sqrt{\sum (x_{ia} - x_{ja})^2}$$

where x_{ia} specifies the position of point i in dimension a. In 'Weighted model', the importance of each dimension is taken in form of weights. It gives the distance between the points i and j as:

$$d_{ij} = \left[\sum_{a} w_{ka} (x_{ia} - x_{ja})^2 \right]^{1/2}$$

where w_{ka} is a diagonal matrix represented the importance of each dimension to each individual.

- *Scaling techniques:* Here we categorize the data in such a way that the properties of original relations remain unchanged. The number of topologies has been proposed but the most long lived classification is that of Stevens (1946). He disguised data in nominal, ordinal, interval and ratio level of measurement that are shown in fig.3. In MDS, mostly ordinal and ratio level are used.

In 'Ordinal level' we categories the data on the basis of ranks Then plotting in n-dimensional space on this rank basis is done. It is further of two types: partial i.e. Some objects while ranking can come at same level, often results from missing data, and complete i.e. no missing data.

In 'Ratio level' Numerical value of distance between objects between adjacent points is consider.

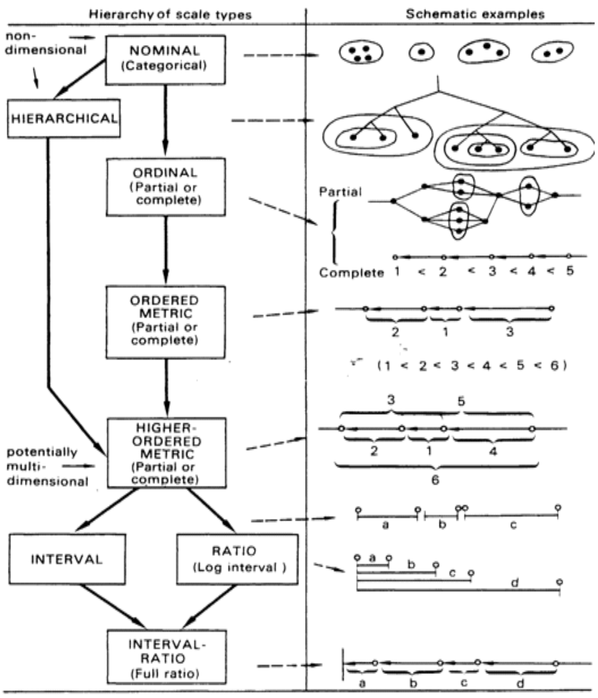

Figure 3: Various Scaling Techniques

III. TYPES OF MDS

MDS can be classified into various types:-

- *Classical MDS*: - It takes only one input matrix giving dissimilarity between pair of items. It uses Euclidean distance model to model the dissimilarities. Two types:-

Metric CMDS:- The data in the matrix represents quantitative similarities. Ratio and interval level of measurements are used. Transformation used is linear transformation i.e.

I{S}= D+E where I{S} is linear transformation of similarities. D is Euclidean distance matrix. E is matrix of error. Data matrix is square and symmetric.

Non metric CMDS: -. The data in matrix represents qualitative similarities. Ordinal level of measurement is used. Transformation is monotonic transformation. i.e.

M{S} = D+E where m{S} is monotonic transformation of similarities. The model is Euclidean distance model. Data is two mode i.e. Matrix is rectangular and asymmetric.

- *Replicated MDS*:- It permits the analyses of several matrices of similarity data simultaneously. It uses same Euclidean distance model as in CMDS but on several input data matrix not on a single matrix. It is also of two types metric and non-metric.
- *Weighted MDS*: In RMDS the difference in responses of individuals is taken into account but in WMDS the cause of difference in response of several individuals is consider To take it into account weight w_k is included that represents the importance of each dimension to each individual. So it is also called individual difference scaling (INDSCAL).

It uses weighted Euclidean distance model. It is also of metric and non metric types. RMDS generates a single distance matrix D, while WMDS generates m unique distance matrices D_k, one for each data matrix S_k.

IV. APPLICATIONS IN LOCALIZATION TECHNIQUES

Applications of MDS are appearing at an increasing rate and at various disciplines such as '*Marketing*' for taking the preferences and perceptions of respondents and '*Data visualizations*' tuning a metric space, in which points correspond to descriptors values, so that the metric space represents a perceptive distance between images. Here main discussion will be on :-

- localization in a wireless sensor network using MDS
- Finding the location of a mobile station in an area of a cellular network using MDS

A. Localization in a wireless sensor network

Determining the physical position of sensor from a large number of sensors randomly and densely deployed in a certain area is important to use the data collected by the sensor. Two approaches are proposed for localization of wireless sensor network:-

a) Centralized approach:- In centralization approach, the inter-node ranging and connectivity data is migrated to a central base station and then the resultinglocations are migrated back to respective nodes. MDS-MAP is centralized algorithm used for network sensor localization that uses

DOI: 01.IJCSI.02.01.42

multidimensional scaling. It uses classic metric MDS.

Let there are n sensors in a network, with positions Xi, i = 1 . . . n, and let matrix X = [X1, X2, . . . ,Xn]^T . X is nxm, where m is the dimensionality of X. Consider m to be an unknown. Let D = [dij] be the nxn matrix of pair-wise distance measurements. The distance measurements dij must obey the triangular inequality: dij + dik >= djk for all (i, j, k).

Classical metric MDS is derived from the 'Law of Cosines', which states that given two sides of a triangle dij , dik, and the angle between them _jik, the third side can be computed using the formula:

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2d_{ij}d_{ik}\cos\theta_{jik}$$

Rewriting:

$$d_{ij}d_{ik}\cos\theta_{jik} = \frac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2)$$

Similarly, $d_{ik} = X_k - X_i$, $d_{jk} = X_k - X_i$. So rewriting we get

$$(X_j - X_i)\cdot(X_k - X_i) = \frac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2)$$

To find the positions X we have to choose some X_0 from X to be the origin of a coordinate system, and construct a matrix B that is known as the matrix of scalar products and the values as follows:

$$b_{ij} = \frac{1}{2}(d_{0i}^2 + d_{0j}^2 - d_{ij}^2)$$

Now, we get a new value of coordinates X'shifted according to origin X_0. Where $X_i' = X_i - X_0$................................ (1)
Now B can also be written as: $X'X'^T = B$
We can solve for X_0 by taking an eigen-decomposition of B into an orthonormal matrix of eigenvectors U and a diagonal matrix of matching eigenvalues V.

$B = X'X'^T = UVU^T$

$X' = UV^{1/2}$ We need to find X in 2-D space or 3D-space. To do this, we throw away all but the two or three largest eigenvalues from V, leaving a 2x2 or 3x3 diagonal matrix, and throw away the matching eigenvectors U. Then X_0 has the proper dimensionality. Now from (1) we can find Xi's. Hence we have determine the coordinates of sensors.

This approach of MDS has some drawbacks as mentioned. A single pass of multidimensional scaling cannot operate on a large topology, MDS MAP also does not use anchor nodes very well as anchor density increases, calculated positions are very near to actual value but don't have exact values, This approach requires distance measurement for all pairs of nodes but some nodes are very far away from a particular sensor and distance can't be measured.

To overcome these difficulties we use another method based on distributed approach.

b) Distributed approach:- This approach is also called hierarchical MDS-based localization algorithm (HMDS).In this architecture, a sensor node may take the *a)* roles of cluster head, cluster member or gateway node. It makes an assumption that whole network is connected i.e. two node have at least one path. Also, each device supports RSSI (Received signal strength indicator measurement). HMDS has three execution phase which are as follows:

1) *Clustering phase:-* Here the whole network is divided into multiple clusters using various algorithms.
2) *Intra cluster localization phase:-* Each cluster contains one cluster head, lots of gateways and many members. Cluster head collects distance measurements from all members of the cluster and compute distances of all pairs of devices using the shortest path algorithm such as the Dijkstra algorithm, so form a distance matrix of cluster members. Applying a MDS on a distance matrix computes the relative coordinates of each cluster member and forms a local map.
3) *Merge phase:-* A gateway node participates in at least two neighbouring clusters, exchanges information between two clusters. Each gateway is assigned at least two coordinates from each cluster the gateway participated in. Different coordinates in the same gateway node implies that the coordinate systems in neighbouring clusters are different, and this information can be used to merge two coordinate systems into a unified coordinate system.

These merged clusters repeatedly broadcast merge requests to neighbouring clusters and calibrate coordinates of the neighbouring clusters until the coordinates of all clusters are consistent and a global map is computed. Hence we get accurate positioning system.

Figure 4: sensor network deployment

B. Finding the location of a mobile station in a cellular network using MDS

Mobile locations estimation can support many applications, such as emergency services, roadside assistance, navigation, tracking etc. Various techniques can be used like GPS, RSS but these suffer from drawbacks like blockage of signal and complex propagation path.

The MDS has been proposed for wireless network based mobile location. Due to the MDS needs distance measurement data between the MS and BSs, the MDS technique has been proposed for mobile location using the TOA (time of arrival) method. Using TOA method, the range can be modelled as: $r_i = d_i + n_i$ where r_i is range measurement between MS and BS_i. n_i is range error that occurs due to multipath fading and propagation losses that includes range delay. To estimate mobile location, a square distance matrix is constructed as:

$$G = \begin{bmatrix} 0 & h_t \\ h_t^T & D_t \end{bmatrix}$$

Where $D_t = \begin{bmatrix} 0 & d_{12}^2 & \cdots & d_{1N}^2 \\ d_{21}^2 & 0 & \cdots & d_{2N}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1}^2 & d_{N2}^2 & \cdots & 0 \end{bmatrix}$ and $h_t = \begin{bmatrix} r_1^2 & r_2^2 & \cdots & r_N^2 \end{bmatrix}^T$ where d_{ij} is

DOI: 01.IJCSI.02.01.42

distance between BS_i and BS_j. h_t is matrix of range between BS and MS. From Dt we can obtain the scalar product matrix Bt that gives the distance of all BSs and MS from a given location that is consider as origin. But distance in matrix Dt is pair wise between either in BSs or BS and MS. Bt is obtained by double centring technique i.e. subtracting rows and columns means from each element. Then add grand mean to each element and multiplying by (-1/2).

When employing the classical MDS, the centroid of the all coordinates is assumed at the origin, so the scalar matrix, Bt can be express as :

$$B_t = XX^T = \begin{bmatrix} X_{MS} \\ X_{bs} \end{bmatrix} \begin{bmatrix} X_{MS} \\ X_{bs} \end{bmatrix}^T \quad \text{............ (2)}$$

Where $X_{bs} = \begin{bmatrix} X_1^T & X_2^T & \dots & X_N^T \end{bmatrix}^T$ matrix of BS with is $X_i = [x_i \quad y_i]$ denotes the coordinates of BS_i.

By eigenvalue decomposition Bt can be represented as :

$B_{T=}UVU^T$

Where V is diagonal matrix of eigenvalues of B_t. And U is a matrix having corresponding eigenvectors.

From this we can evaluate the value of X as

$X= UV^{1/2}$

Now if we know the value of X, also all BSs are fixed and hence we know the value of coordinates of BSs. Thus, from (2), we can find the value of X_{MS} i.e. coordinates of MS.

Classical multidimensional scaling is sensitive to the range measurements error in the NLOS (non line of sight) scenario.

Conclusion

This paper gives an overview of multidimensional scaling algorithm and its current practical applications. It is an optimization algorithm that has very bright scope in future and will help in understanding and solving large complex processes which seems to be very difficult in today's scenario.

References

[1] Multidimensional Scaling, by Forrest W.Young, University ofNorth Californiahttp://forrest.psych.unc.edu/teaching/p208a/mds/mds.html.

[2] Borg, I. and Groenen, P.: "Modern Multidimensional Scaling: theory and applications" (2nd ed.), Springer-Verlag New York, 2005.

[3] Cox, T.F., Cox, M.A.A., (2001), Multidimensional Scaling, Chapman and Hall.

[4] P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg, I. Olkin, N. Wermuth, S. Zeger, "Modern Multidimensional Scaling Theory and Applications," Springer, 1997.

[5] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102-114, Aug 2002.

[6] G. B. J. Hightower, "Locations systems for ubiquitous computing," *IEEE Computer*, vol. 34, pp. 57–66, 2001.

[7] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from connectivity in sensor networks," *IEEE Trans. Parallel Distrib.Syst.*, vol. 15, no. 11, pp. 961–974, 2004.

[8] X. Ji and H. Zha, "Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling," in *Proc. IEEE INFOCOM*, March 2004.

[9] Sau Yee Wong, Joo Ghee Lim, SV Rao and Winston KG Seah, "Density-aware hop-count localization (DHL) in wireless sensor networks with variable density," *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC2005)*, vol. 3, pp. 1848-1853, Mar. 2005.

[10] P. Biswas and Y. Ye, "Semidefinite programming for ad hoc wireless sensor network localization," in *IPSN '04: Proceedings of the 3rd international symposium on Information processing in sensor networks*. New York, NY, USA: ACM, 2004, pp. 46–54.

[11] P. Drineas, A. Javed, M. Magdon-Ismail, G. Pandurangant, R. Virrankoski, and A. Savvides, "Distance matrix reconstruction from incomplete distance information for sensor network localization," in *Proceedings of Sensor and Ad-Hoc Communications and Networks Conference (SECON)*, vol. 2, Sept. 2006, pp. 536–544.

[12] Gwo-Jong Yu and Shao Wang "A Hierarchical MDS-based Localization Algorithm for Wireless Sensor Networks",Journal of Information Technology and Applications.

[13] C. Drane, M. Macnaughtan, and C. Scott, "Positioning GSM telephones," *IEEE Commun. Mag.*, vol. 36, no. 4, pp. 46–55, Apr. 1998

[14] E. Kapplan, "Understanding GPS: Principles and Applications," *Artech House*, 1996.

[15] G.D. Ott, "Vehicle Location in Cellular Mobile Radio Systems", *IEEE Trans. on Veh. Technol.*, vol. VT-26, pp. 43-46, Feb. 1977.

[16] J. C. Liberti Jr. and T. S. Rappaport, "Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications,"*Prentice Hall*, 1999.

[17] M. Pent, M.A. Spirito, and E. Turco, "Method for positioning GSM mobile stations using absolute time delay measurements," E*lectronics Letters*, vol. 33, pp. 2019-2020, Nov. 1997.

[18] K. W. Cheung, H. C. So, "A multidimensional scaling framework for mobile location using time-of-arrival measurements," *IEEE Trans. Signal Process.*, vol. 53, pp. 460 – 470, Feb 2005.

[19] R.-T. Juang, D.-B. Lin, and H.-P. Lin, "Hybrid SADOA/TDOA mobile positioning for cellular networks," *IET Communications*, vol. 1, Issue 2,pp. 282-287, Apr. 2007.

DOI: 01.IJCSI.02.01.42